PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F

A2

(11) International Publication Number: WO 99/34271

(43) International Publication Date: 8 July 1999 (08.07.99)

(21) International Application Number: PCT/IB98/02090

(22) International Filing Date: 21 December 1998 (21.12.98)

(30) Priority Data:
97204130.5 29 December 1997 (29.12.97) EP

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(71) Applicant *(for SE only)*: PHILIPS AB [SE/SE]; Kottbygatan 7, Kista, S-164 85 Stockholm (SE).

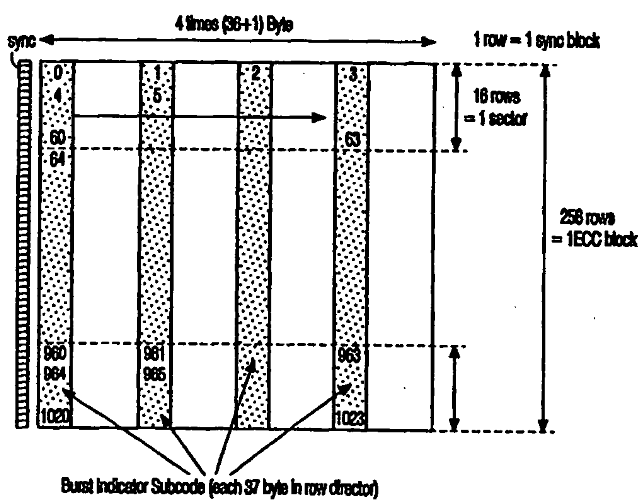
(72) Inventors: TOLHUIZEN, Ludovicus, M., G., M.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). VAN DIJK, Marten, E.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). BAGGEN, Constant, P., M., J.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: VAN DER KRUK, Willem, L.; Internationaal Octrooibureau B.V., P.O. Box 220, NL-5600 AE Eindhoven (NL).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: A METHOD FOR ENCODING MULTIWORD INFORMATION BY WORDWISE INTERLEAVING AND ERROR PROTECTION, WITH ERROR LOCATIVE CLUES DERIVED FROM HIGH PROTECTIVITY WORDS AND DIRECTED TO LOW PROTECTIVITY WORDS, A METHOD FOR DECODING SUCH INFORMATION, A DEVICE FOR ENCODING AND/OR DECODING SUCH INFORMATION, AND A CARRIER PROVIDED WITH SUCH IN



(57) Abstract

Multiword information is encoded as based on multibit symbols in relative contiguity with respect to a medium, whilst providing wordwise interleaving and wordwise error protection code facilities. This may provide error locative clues across multiword groups, that originate in high protectivity clue words and are directed to low protectivity target words. Further, the clue words may have a first uniform size and be interspersed in a first uniform manner. The target words may have a second uniform size and be interspersed in a second uniform manner. In particular, the organization may be applied for use with optical storage.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

A method for encoding multiword information by wordwise interleaving and error protection, with error locative clues derived from high protectivity words and directed to low protectivity words, a method for decoding such information, a device for encoding and/or decoding such information, and a carrier provided with such information.

BACKGROUND OF THE INVENTION

The invention relates to a method as recited in the preamble of Claim 1. US Patents 4,559,625 to Berlekamp et al, and US 5,299,208 to Blaum et al disclose the decoding of interleaved and error protected information words, wherein an error pattern found in a first word may give a clue to locate errors in another word of the same group of words. The references use a standardized format and a fault model that has multisymbol error bursts across various words. Occurrence of an error in a particular word presents a strong probability for an error to occur at a corresponding symbol position pointed at in a next word or words. This procedure often raises the number of corrected errors. The present inventors have recognized a problem with this principle: a clue will only materialize when the clue word has been fully corrected.

SUMMARY TO THE INVENTION

In consequence, amongst other things, it is an object of the present invention to provide a coding format wherein clue words will be correctly decoded with a greater degree of certainty than a target word. Now therefore, according to one of its aspects the invention is characterized according to the characterizing part of Claim 1. The clue found may result in or point to an **erasure** symbol. With such pointing, error correction will proceed in a more powerful manner. In fact, many codes will correct at most t errors when no error locative indication is known. Given the erasures locations, generally a larger number $e>t$ of erasures may be corrected. Also, the protection against a combination of bursts and random errors will improve. Alternatively, the providing of erasure locations will require the use of only a lower number of syndrome symbols, thus simplifying the calculation. In principle, the invention may be used in a storage environment as well as in a transmission environment.

The invention also relates to a method for decoding information so encoded, to an encoding and/or decoding device for use with the above method, and to a carrier provided with information for interfacing to such encoding and/or decoding. Further advantageous aspects of the invention are recited in dependent Claims.

BRIEF DESCRIPTION OF THE DRAWING

These and further aspects and advantages of the invention will be discussed more in detail hereinafter with reference to the disclosure of preferred embodiments, and in particular with reference to the appended Figures that show:

Figure 1, a system with encoder, carrier, and decoder;

Figure 2, a code format principle;

Figure 3, a product code format;

Figure 4, a Long Distance Code with burst detection;

Figure 5, a picket code and burst indicator subcode;

Figure 6, a burst indicator subcode format;

Figure 7, a picket code and its product subcode;

Figure 8, various further aspects thereof;

Figure 9, an alternative format;

Figure 10, a detail on the interleaving.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Figure 1 shows a comprehensive system according to the invention, that is provided with an encoder, a carrier, and a decoder. The embodiment is used for encoding, storing, and finally decoding a sequence of samples or multibit symbols derived from an audio or video signal, or from data. Terminal 20 receives a stream of symbols that by way of example have an eight bit size. Splitter 22 recurrently and cyclically transfers first symbols intended for the clue words to encoder 24. Furthermore, splitter 22 transfers all other symbols to encoder 26. In encoder 24 the clue words are formed by encoding the associated data into code words of a first multi-symbol error correcting code. This code may be a Reed-Solomon code, a product code, an interleaved code, or a combination thereof. In encoder 26 the target words are formed by encoding into code words of a second multi-symbol error correcting code. In the embodiment, all code words will have a uniform length, but this is not a strict requirement. Preferably, both codes will be Reed-Solomon codes with the first one a subcode of the second code. As will become more clear with respect to Figure 2, the clue words will have in general a much higher degree of error protection, and contain relatively fewer non-redundant symbols.

In block 28, the code words so formed are transferred to one or more outputs of which an arbitrary number has been indicated, so that the distribution on a medium to be discussed later will become uniform. Block 30 symbolizes the medium itself

that receives the encoded data. This may in fact relate to direct writing in an appropriate write-mechanism-plus-medium combination. Alternatively, the medium may be realized as a copy from a master encoded medium such as a stamp. Preferably, storage will be optical and fully serial, but other configurations may be used. In block 32, the various words will be read again from the medium. Then the clue words of the first code will be sent to decoder 34, and decoded as based on their inherent redundancies. Furthermore, as will become apparent in the discussion of Figure 2 hereinafter, such decoding may present clues on the locations of errors in other than these clue words. Box 35 receives these clues and contains a program for using one or more different strategies for translating such clues to erasure locations. The target words are decoded in decoder 36. Under control of the erasure locations, the error protection of the target words is raised to an acceptable level. Finally, all decoded words are demultiplexed by means of element 38 conformingly to the original format to output 40. For brevity, the mechanical configuration for interfacing the various subsystems to each other have been omitted.

Figure 2 illustrates a relatively simple code format. As shown, the coded information has been notionally arranged in a block of 16 rows and 32 columns of symbols, that is 512 symbols. Storage on a medium is serially column-by-column starting at the top left column. The hatched region contains check symbols, and words 0, 4, 8, and 12 have 8 check symbols each and constitute clue words. The other words contain 4 check symbols each and constitute target words. The whole block contains 432 information symbols and 80 check symbols. The latter may be localized in a more distributed manner over their respective words. A part of the information symbols may be dummy symbols. The Reed-Solomon code allows to correct in each clue word up to four symbol errors. Actual symbol errors have been indicated by a crosses. In consequence, all clue words may be decoded correctly, inasmuch as they never have more than four errors. Notably words 2 and 3 may however not be decoded on the basis of their own redundant symbols only. Now, in Figure 2 all errors, except 62, 66, 68 represent error strings. However, only strings 52 and 58 that cross at least three consecutive clue words, are considered as error bursts, so that at least all intermediate symbol locations get an erasure flag. Also, the target words before the first clue word error of the burst and the targets words just after the last clue word error of the burst may get an erasure flag at that location, depending on the strategy followed. String 54 is not considered a burst, because it is too short.

As a consequence, the two errors in word 4 produce an erasure flag in both associated columns. This renders words 2 and 3 correctable, each with a single error

symbol and two erasure symbols. However, random errors 62, 68, nor string 54 constitute clues for words 5, 6, 7, because each of them contains only a single clue word. In certain situations, an erasure may result in a **zero** error pattern, because an arbitrary error in an 8-bit symbol has a 1/256 probability to cause again a correct symbol. Likewise, a long burst crossing a particular clue word may produce a correct symbol therein. By a bridging strategy between preceding and succeeding clue symbols of the same burst, this correct symbol is then incorporated into the burst, and in the same manner as erroneous clue symbols translated into erasure values for appropriate target symbols. The above decisions may be amended according to decoding policy, that may further be controlled by other parameters.

DISCUSSION OF A PRACTICAL FORMAT

Hereinafter, a practical format will be discussed. Figure 3 symbolizes a product code format. Words are horizontal and vertical, and parity is hatched. Figure 4 symbolizes a so-called Long Distance Code with special burst detection in the upper few words that have more parity. The invention presents a so-called **Picket Code** that may be constructed as a combination of the principles of Figures 3 and 4. Always, writing is sequential along the arrows shown in Figures 3, 4.

Practical aspects of the present invention are brought about by newer methods for digital optical storage. A particular feature is that in the case of substrate incident reading the upper transmissive layer is as thin as 100 micron. The channel bits have a size of some 0.14 micron, so that a data byte at channel rate of 2/3 will have a length of only 1.7 microns. The beam diameter at the top surface has a diameter of some 125 microns. A caddy or envelope for the disc will reduce the probability of large bursts. However, non-conforming particles of less than 50 microns may cause short faults The inventors have inter alia used a fault model wherein such faults through error propagation may lead to bursts of 200 microns, corresponding to some 120 Bytes. In particular, the inventors have used an error model with fixed size bursts of 120B that start randomly with a probability per byte of $2.6*10^{-5}$, or on the average one burst per 32kB block. The invention has been pushed by developments in optical disc storage, but other configurations such as multitrack tape, and other technologies such as magnetic and magneto-optical would also benefit from the improved approach described herein.

Figure 5 shows a picket code and burst indicator subcode. A picket code consists of two subcodes A and B. The burst indicator subcode (BIS) contains the clue words. By format, it is a very deeply interleaved long distance code that allows to localize

the positions of the multiple burst errors. The error patterns so found are processed to obtain erasure information for the target words that are configured in this embodiment as a product subcode (PS). The product subcode will correct combinations of multiple bursts and random errors, through using erasure flags obtained from the burst indicator subcode.

The following format is proposed:

- the block of '32 kB' contains 16 DVD-compatible sectors
- each such sector contains 2064=2048+16 Bytes data
- each sector after ECC encoding contains 2368 Bytes
- therefore, the coding rate is 0.872
- in the block, 256 sync blocks are formatted as follows
- each sector contains 16 sync blocks
- each sync block consists of 4 groups of 37 B
- each group of 37 B contains 1 B of deeply interleaved Burst Indicator Subcode and further 36 B of Product Subcode.

As shown in Figure 5, rows are read sequentially from the disc, starting with the preceding sync pattern. Each row contains 4 B of the BIS shown in hatched manner and numbered consecutively, and separated by 36 other bytes. Sixteen rows form one sector and 256 rows form one sync block.

Figure 6 shows exclusively a burst indicator subcode format of the same 64 numbered bytes per sector of Figure 5, and is constructed as follows:

- there are 16 rows, with each a [64,32,33] RS code with t=16;
- columns derive in sequence from disk as shown by the arrow, so that groups of four columns derive from a single sector for fast addressing;
- BIS may indicate at least 16 bursts of 592 B (~1 mm.) each;
- BIS contains 32 B data per sector, 4 columns of the BIS, and in particular 16 B DVD header, 5 B parity on the header to allow fast address readout and 11 B user data.

Figure 7 shows a picket code and its product subcode that is built from the target words. The Bytes of the Product Subcode are numbered in the order as they are read from the disc, whilst ignoring the BIS bytes.

Figure 8 shows various further aspects of the of this embodiment of the product subcode. In particular, the product subcode is a [256,228,29]*[144,143,2] Product Code of Reed-Solomon codes. The number of data Bytes is 228*143=32604, that is sixteen times (2048+11) user Bytes plus 12 spare Bytes.

Figure 9 shows an alternative format to Figure 8, leaving out the

horizontal Reed-Solomon code altogether. The horizontal block size is 36 bytes (one quarter of Figure 7), and uses a [256,224,33] Reed-Solomon code. Each sector has 2368 Bytes and no dummy Bytes are necessary.

The code in the first column is formed in two steps. From each sector, the 16 header Bytes are encoded in a [20,16,5] code first to allow fast address retrieving. The resulting 20 Bytes plus a further 32 user bytes per sector form data bytes and are collectively encoded further. The data symbols of one 2K sector may lie in only one physical sector, as follows. Each column of the [256,224,33] code contains 8 parity symbols per 2k sector. Further, each [256,208,49] code has 12 parity symbols per 2K sector and 4 parity symbols of the [20,16,5] code to get a [256,208,49] code with 48 redundant bytes.

Figure 10 shows this interleaving in detail. Here, '*' represents the header Bytes, '□' the parities of the [20,16] code, '•' the 32 "further" data bytes and 12 parity bytes for the [256,208] code.

CLAIMS:

1. A method for encoding multiword information that is based on multibit symbols in relative contiguity with respect to a medium, whilst providing wordwise interleaving and wordwise error protection code facilities, for so providing error locative clues across multiword groups,

characterized by originating such clues in high protectivity clue words as being directed to low protectivity target words.

2. A method as claimed in Claim 1, wherein such clue words have a first uniform size and are interspersed in a first uniform manner with respect to target words that have a second uniform size and are interspersed in a second uniform manner.

3. A method as claimed in Claim 1, and applied to storage with respect to an optical medium.

4. A method as claimed in Claim 1, wherein said clue words contain header informations of associated sectors within a block that contains the above code facilities, and which header informations are presented to said medium in sequential correspondence to a disposition of the respectively associated sectors.

5. A method as claimed in Claim 4, wherein per sector the header informations have additional error protection outside said code facilities.

6. A method for decoding received multiword information that is based on multibit symbols presented in relative contiguity with respect to a medium, whilst effecting wordwise de-interleaving and decoding of error protection code facilities, inclusive of evaluating error locative clues across multiword groups,

characterized by deriving such clues from high protectivity clue words as being directed to low protectivity target words.

7. A method as claimed in Claim 6, being based on clue words that have a first uniform size, and are interspersed in a first uniform manner and target words that have a second uniform size, and are interspersed in a second uniform manner.

8. A method as claimed in Claim 6 and applied to storage with respect to an optical medium.

9. A method as claimed in Claim 6, wherein corrected symbols in clue

words provide respective clues, and successive clues in a series of received informations collectively yield erasure flags for intermediate symbols of target words.

10. A method as claimed in Claim 9, wherein to an intermediate unaltered clue word symbol within such series a notional clue is assigned.

11. A method as claimed in Claim 6, wherein said clue words contain header informations of associated sectors within a block that contains the above code facilities, and deriving said header informations from said medium in sequential correspondence to a disposition of the respectively associated sectors.

12. A method as claimed in Claim 11, whilst undertaking error protection per sector on the header informations outside said code facilities.

13. A device for encoding multiword information that is based on multibit symbols in relative contiguity with respect to a medium, and having interleave means for providing wordwise interleaving, coding means for providing wordwise error protection code facilities, and assign means for producing error locative clues across multiword groups, wherein such assign means are arranged to provide such clues to originate in high protectivity clue words and point to low protectivity target words.

14. A device as claimed in Claim 13, wherein such interleave means are arranged for interleaving such clue words at a first uniform size as being interspersed in a first uniform manner with respect to target words that have a second uniform size and are interspersed in a second uniform manner.

15. A device for decoding received multiword information that is based on multibit symbols presented in relative contiguity with respect to a medium, and having de-interleave means for effecting wordwise de-interleaving and decoding means for decoding error protection code facilities, and evaluating means for evaluating error locative clues across multiword groups,

characterized in that said evaluating means are arranged for deriving such clues from high protectivity clue words as being directed to low protectivity target words.

16. A device as claimed in Claim 15, being based on clue words that have a first uniform size, and are interspersed in a first uniform manner, and target words that have a second uniform size, and are interspersed in a second uniform manner.

17. A physical carrier produced by practising a method as claimed in Claim 1, comprising an array of interleaved clue words and target words, such clue words having a superior error protectivity with respect to said target words.

18. A carrier as claimed in Claim 17, wherein such clue words have a first

uniform size and are interspersed in a first uniform manner with respect to target words that have a second uniform size and are interspersed in a second uniform manner.

19. A carrier as claimed in Claim 17, and being based on optical storage.

20. A carrier as claimed in Claim 17, and being intended for use with substrate incident reading.

21. A carrier as claimed in Claim 17, wherein said clue words contain header informations of associated sectors within a block that contains said code facilities, and which are presented to said medium in sequential correspondence to a disposition of the respectively associated sectors.

22. A carrier as claimed in Claim 21, wherein per sector the header informations have error protection outside said code facilities.

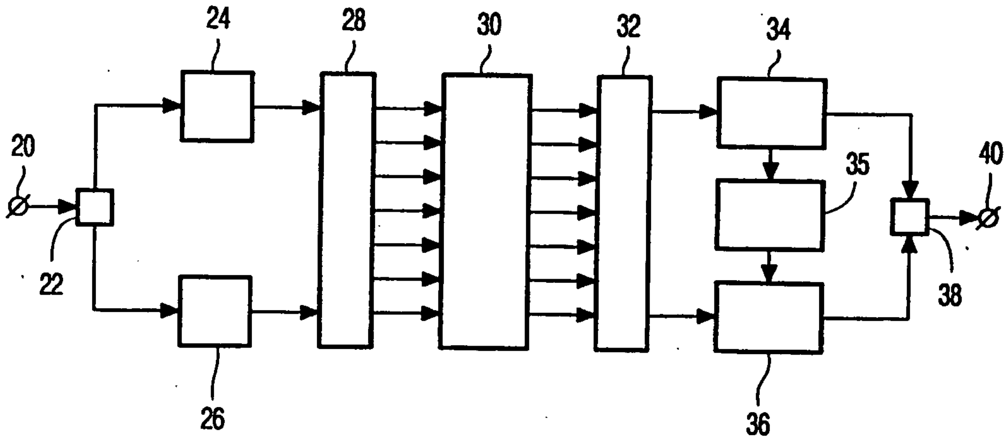

20
22
24
26
28
30
32
34
35
36
38
40

FIG. 1

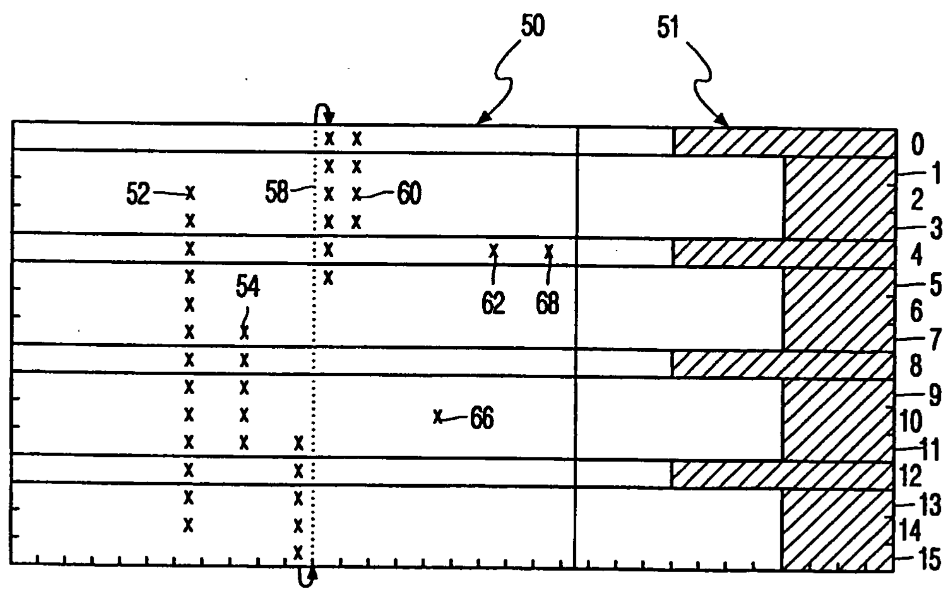

50
51
52
54
58
60
62
68
66
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

FIG. 2

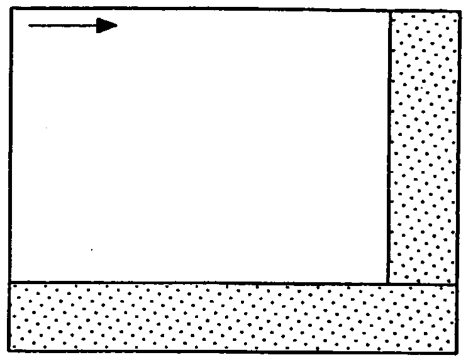


FIG. 3

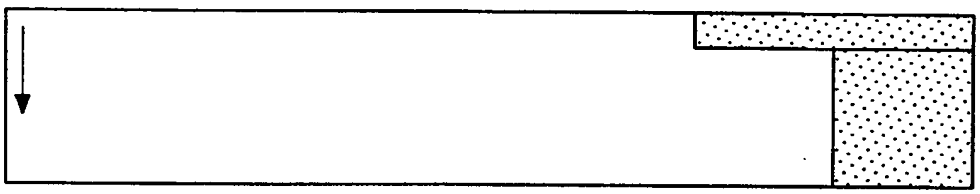


FIG. 4

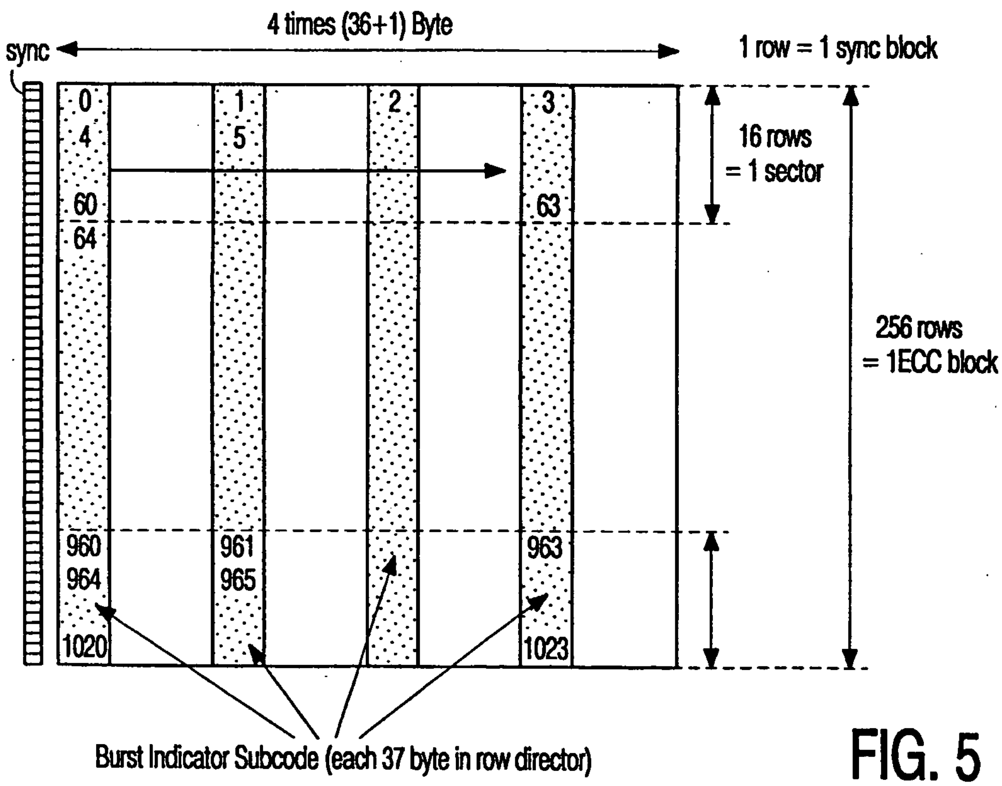
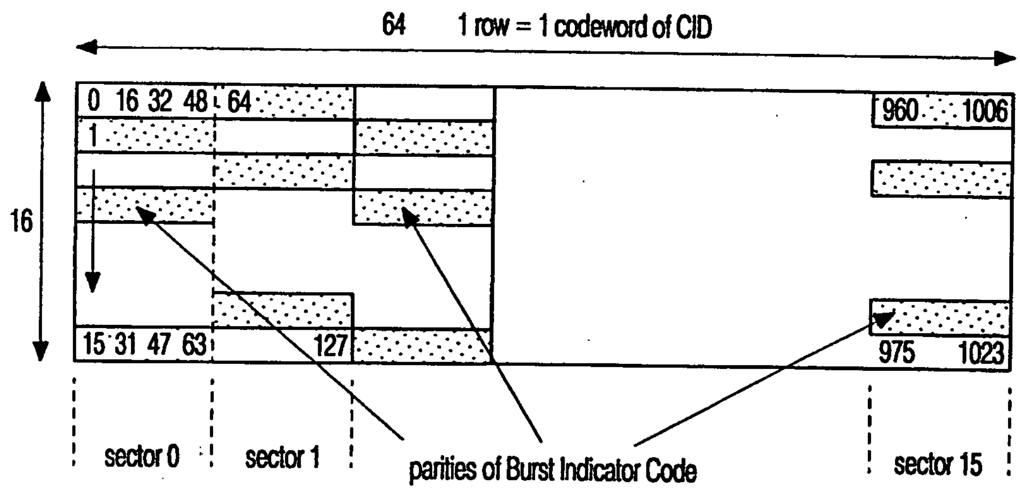
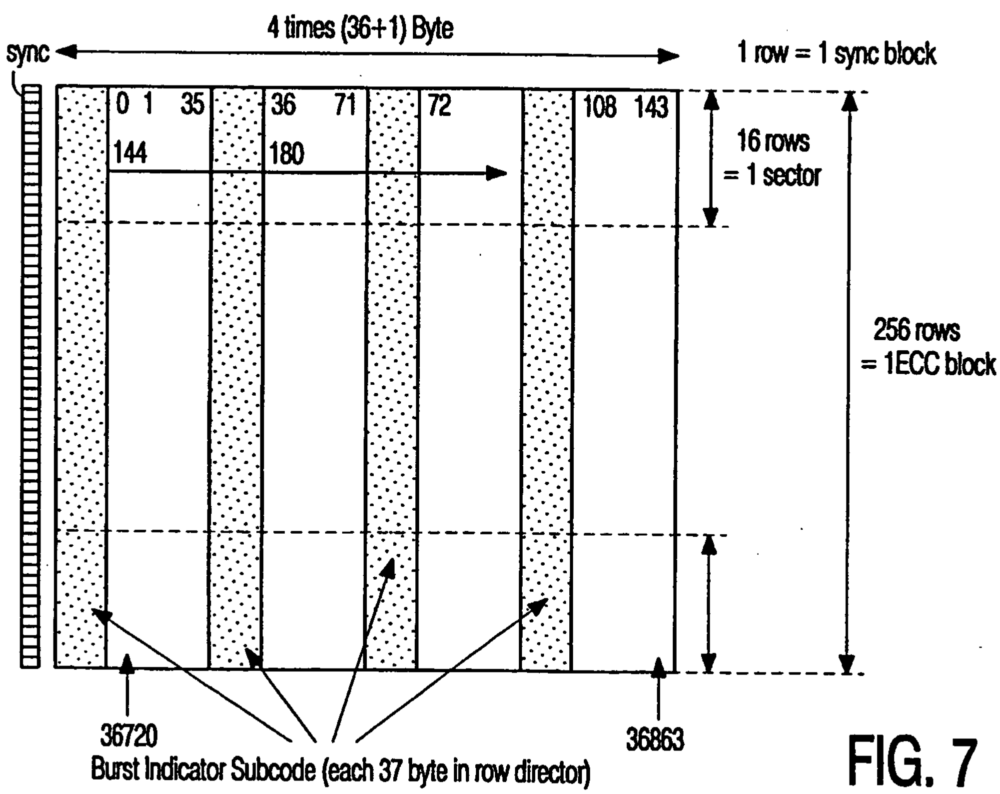


FIG. 5



64
1 row = 1 codeword of CID
16
0 16 32 48 64
1
960 1006
15 31 47 63
127
975 1023
sector 0
sector 1
parities of Burst Indicator Code
sector 15

FIG. 6



4 times (36+1) Byte
sync
1 row = 1 sync block
0 1 35
36 71
72
108 143
144
180
16 rows
= 1 sector
256 rows
= 1ECC block
36720
36863
Burst Indicator Subcode (each 37 byte in row director)

FIG. 7

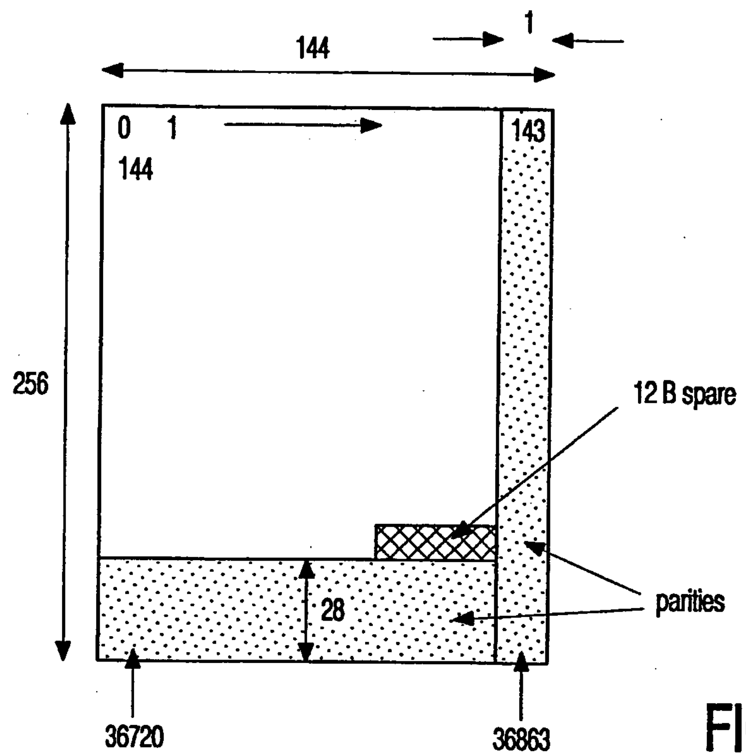


FIG. 8

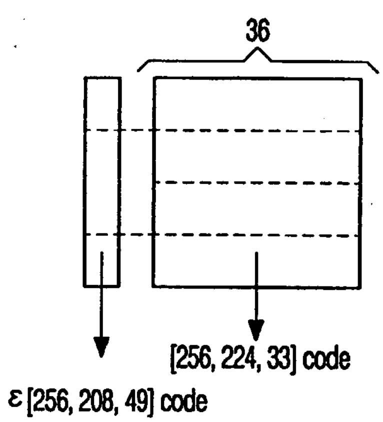


FIG. 9

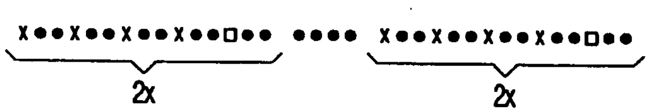


FIG. 10